

where e - vaccine efficacy;
 v - vaccination rate.

A certain amount of susceptible will be vaccinated and move into the recovered class without becoming infected.

Conclusions. An analysis of SIR mathematical model of the spread of infectious diseases and its modifications is made in this article. In particular, SEIR model is considered and its features are described. An epidemiological characteristics of measles are studied. An appropriate model for modeling the spread of measles is selected and improved by adding two parameters: vaccination and vaccine efficacy.

REFERENCES:

1. Official site of the Public Institution of the Center of Public Health of the Ministry of Health of Ukraine [Electronic resource] <https://phc.org.ua/news/dani-zakhvoryuvanosti-na-kir-10-16-zhovtnya-2019>
2. World Health Organization [Electronic resource] <https://www.who.int/news-room/fact-sheets/detail/measles>
3. Kermack W.O., McKendrick A.G., A Contribution to the Mathematical Theory of Epidemics [Text] // Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 115 (772), 1927.

УДК 004.9

ВИКОРИСТАННЯ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ДЛЯ ДОСЛІДЖЕННЯ МОДЕЛІ ВИБОРЧОЇ КОМПАНІЇ

В.В. Гнатушенко, О.Г. Гончаров
(Україна, Дніпро, Національна металургійна академія України)

Постановка проблеми. Дослідження аналітичних методів Microsoft SQL в галузі інтелектуального аналізу даних для створення моделі виборчої компанії. Для цього необхідно виділити підмножини вирішуваних завдань, провести порівняльний аналіз методів інтелектуального аналізу даних для вибраних типів завдань і на практичному прикладі визначити випадки, для яких більш привабливим є той чи інший метод.

Різноманіття методів для вирішення цього завдання ставить перед аналітиком питання вибору алгоритму, який найкращим чином придатний підходить під поставлені вимоги задачі [1, 2]. Так, у разі кластеризації, рекомендується використовувати ієрархічні методи, якщо заздалегідь невідомо число кластерів і потрібно отримати детальне уявлення про структуру даних. У свою чергу, ітеративні методи характеризуються більш високою стійкістю по відношенню до шумів і викидів, некоректного вибору метрики, включенню в аналіз незначущих атрибутів, але вимагають апріорної вказівки числа кластерів. У разі вибору методу класифікації необхідно брати до уваги такі параметри, як

точність, інтерпретуємість результатів і, в залежності від випадку, масштабованість.

В рамках даної роботи було реалізовано практичне застосування методів ІАД в середовище MS SQL Server. Як об'єкт дослідження були взяті президентські праймеріз США 2016 року. Після проведення аналізу та формалізації предметної області визначені завдання дослідження (кластеризація та класифікація) і зроблена їх формальна постановка.

Проведена кластеризація громадян, які голосували на демократичних і республіканських праймеріз, тобто тих, хто в переважній більшості випадків відносить себе до відповідної партії. Кластеризація проводилася на двох незалежних вибірках: серед тих регіонів, де встигли пройти республіканські праймеріз, і тих, де пройшли демократичні. Для республіканців був складений портрет «типового» виборця, що віддає свій голос за Дональда Трампа. Однак, якщо поглянути на середні значення виділених ознак по всіх округах, можна встановити, що отриманий портрет збігається з портретом «середнього американця». Це говорить про те, що використані алгоритми не змогли виявити демографічних особливостей груп виборців, які віддають перевагу Дональду Трампу та могли б пояснити його успіх.

Проведено порівняльний аналіз двох алгоритмів, які засновані на методі максимізації очікувань і методі К-середніх. Було встановлено, що алгоритм максимізації очікувань в нашому випадку показує кращі результати, ніж алгоритм К-середніх. Отриманий результат можна пояснити тим, що алгоритм К-середніх має перевагу при роботі з відокремленими кластерами, але значно поступається алгоритму максимізації очікування при наявності їх перекриття, а саме цей випадок характеризує початковий набір даних.

Побудовано моделі класифікації. Досліджено вплив параметрів, що настроюються на чутливість одержуваних моделей і точність прогнозування. Завданням було підібрати найкращий алгоритм, який визначає ймовірність перемоги одного з демократів на своїх праймеріз. Були розглянуті методи дерева рішень, нейронна мережа і байесовський. Найвищу точність показала модель на основі нейронної мережі. Крім цього, вдалося підібрати такі параметри алгоритму, які підвищили її якість в порівнянні з настройками за замовчуванням. Отримана модель була використана для прогнозування результату демократичних праймеріз в штатах, де попередні голосування проходили з квітня по червень включно.

Висновки. Виконані всі етапи інтелектуального аналізу даних починаючи від аналізу предметної області і закінчуючи застосуванням моделі для реальної задачі прогнозування виборчої компанії.

ПЕРЕЛІК ПОСИЛАНЬ:

1. Н.Б. Паклин, В.И.Орешков. Бизнес-аналитика: от данных к знаниям.– СПб: Питер, 2009.– 624с.
2. Замятин А.В. Интеллектуальный анализ данных Учебное пособие. - Томск: Издательский Дом Томского государственного университета, 2016. - 120 с. - ISBN 978-5-94621-531-2